

APPENDIX L

Input: 1. A dataset consist of continuous and dummy variables that it is normalized, X;
 2. A target variable y.
 Output: 1. Selected Principle Components W
 2. Corresponding loading matrix U
 3. success //a flag indicating whether SVD successful: 0; or not:0
 Parameter : Percentage variance to keep AE. Default 0.9. Range : 0.8~0.95.
 Process:
 $NE = X^T * X;$ //NE is the normal equation matrix
 $[U \ S \ V] = svd(NE);$ //use svdcmp function from Numeric Recipe
 If SVD succeeds
 success = 0
 Else
 success = -1,
 W and U both empty
 End If
 Sort the singular values in S in descending order;
 Re-arrange columns in U, make them still correspond to their singular values;
 Set n = the number of columns in X;
 enough_e = n * AE;
 sume = 0;
 TU = empty; TW = empty;
 i = 1;
 While (sume < enough_e and S(i,i) > 0.1)
 TU = [TU, U(:,i)]; //U(:,i) is the ith column of U
 TW = [TW, W(:,i)]; // W(:,i) is the ith column of W
 sume += S(i,i);
 i++;
 End While;
 While (S(i,i) > 0.1)
 corr = absolute value of correlation of W(:,i) and y;
 If (corr > 0.3)
 TU = [TU, U(:,i)];
 TW = [TW, W(:,i)];
 End If
 i++;
 End While
 U = TU; W = TW;
 Return W, U, success.